The Self
The Self
The Self
The Self
The Self
The Self
The Self
THE SELF
THE SELF

The Self

A HISTORY

OXFORD PHILOSOPHICAL CONCEPTS

Edited by PATRICIA KITCHER

The Self

OXFORD PHILOSOPHICAL CONCEPTS

OXFORD PHILOSOPHICAL CONCEPTS

Christia Mercer, Columbia University

Series Editor

PUBLISHED IN THE OXFORD PHILOSOPHICAL CONCEPTS SERIES

Efficient Causation
Edited by Tad Schmaltz

Sympathy
Edited by Eric Schliesser

The Faculties
Edited by Dominik Perler

Memory
Edited by Dmitri Nikulin

Moral Motivation
Edited by Iakovos Vasiliou

Eternity
Edited by Yitzhak Melamed

Self-Knowledge
Edited by Ursula Renz

Embodiment
Edited by Justin E. H. Smith

Dignity
Edited by Remy Debes

Animals
Edited by G. Fay Edwards and Peter Adamson

Pleasure
Edited by Lisa Shapiro

Health
Edited by Peter Adamson

Evil
Edited by Andrew Chignell

Persons
Edited by Antonia LoLordo

Space
Edited by Andrew Janiak

Teleology
Edited by Jeffrey K. McDonough

The World Soul
Edited by James Wilberding

Powers
Edited by Julia Jorati

The Self
Edited by Patricia Kitcher

FORTHCOMING IN THE OXFORD PHILOSOPHICAL CONCEPTS SERIES

Modality
Edited by Yitzhak Melamed

Human
Edited by Karolina Hubner

Love
Edited by Ryan Hanley

The Principle of Sufficient Reason
Edited by Fatema Amijee and Michael Della Rocca

OXFORD PHILOSOPHICAL CONCEPTS

The Self

A HISTORY

Edited by Patricia Kitcher

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

© Oxford University Press 2021

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data
Names: Kitcher, Patricia, editor.
Title: The self : a history / edited by Patricia Kitcher.
Other titles: Self (Oxford University Press)
Description: New York : Oxford University Press, [2021] |
Includes bibliographical references and index.
Identifiers: LCCN 2020058296 (print) | LCCN 2020058297 (ebook) |
ISBN 9780190087265 (hardback) | ISBN 9780190087258 (paperback) |
ISBN 9780190087289 (epub)
Subjects: LCSH: Self (Philosophy)
Classification: LCC BD438.5 .S4457 2021 (print) | LCC BD438.5 (ebook) |
DDC 126—dc23
LC record available at https://lccn.loc.gov/2020058296
LC ebook record available at https://lccn.loc.gov/2020058297

DOI: 10.1093/oso/9780190087265.001.0001

1 3 5 7 9 8 6 4 2

Paperback Printed by Marquis, Canada
Hardback printed by Bridgeport National Bindery, Inc., United States of America

Contents

SERIES EDITOR'S FOREWORD vii

ACKNOWLEDGMENTS ix

CONTRIBUTORS xi

Introduction—Self: History of a Concept 1
PATRICIA KITCHER

1 Augustine on Cogitation and Self-Constitution: Drawing from and Surpassing the Plotinian Stance 28
PAULIINA REMES

2 The Heritage of Ibn Sīnā's Concept of the Self 55
JARI KAUKUA

3 Aquinas and "I": A Medieval Concept of Self 73
THERESE SCARPELLI CORY

4 Descartes on Subjects and Selves 99
VILI LÄHTEENMÄKI

5 Locke on Being Self to My Self 118
RUTH BOEKER

6 Leibniz on the Idea and Concept of the Self : Inner Sentiment, Self-Knowledge, and the Awakening of Metaphysical Ideas 145
CHRISTIAN BARTH

Reflection: Metacognition and the Self-Concept 172
JANET METCALFE

7 The Idea of Self in Hume's *Treatise* 179
DON GARRETT

8 Diderot, the Self, and the Science of Dreaming 212
CHARLY COLEMAN

Reflection: Caravaggio's Self-Regard 233
MARIA H. LOH

9 Kant on the Unity of Self-Consciousness and Moral Agency 240
PATRICIA KITCHER

10 Self-Awareness and the "I" in the Phenomenological Tradition 267
SACHA GOLOB

Reflection: "You Are an *I*": Elizabeth Bishop's "In the Waiting Room" 287
GARY OSTERTAG

11 Self and Self-Representation in the Long Twentieth Century: A Critical Discussion 295
CHRISTOPHER PEACOCKE

Reflection: Spontaneous Neural Activity and the Self: A Neuroscience Perspective 317
ALISON HANSON AND RAFAEL YUSTE

12 Rethinking the Self within an African Philosophical Paradigm 326
ANTJIE KROG

BIBLIOGRAPHY 349

INDEX 381

Series Editor's Foreword

Oxford Philosophical Concepts (OPC) offers an innovative approach to philosophy's past and its relation to other disciplines. As a series, it is unique in exploring the transformations of central philosophical concepts from their ancient sources to their modern use.

OPC has several goals: to make it easier for historians to contextualize key concepts in the history of philosophy, to render that history accessible to a wide audience, and to enliven contemporary discussions by displaying the rich and varied sources of philosophical concepts still in use today. The means to these goals are simple enough: eminent scholars come together to rethink a central concept in philosophy's past. The point of this rethinking is not to offer a broad overview, but to identify problems the concept was originally supposed to solve and investigate how approaches to them shifted over time, sometimes radically. Recent scholarship has made evident the benefits of reexamining the standard narratives about western philosophy. OPC's editors look beyond the canon and explore their concepts over a wide philosophical landscape. Each volume traces a notion from its inception as a solution to specific problems through its historical transformations to its modern use, all the while acknowledging its historical context. Each OPC volume is a history of its concept in that it tells a story about changing solutions to its well-defined problem. Many editors have found it appropriate

to include long-ignored writings drawn from the Islamic and Jewish traditions and the philosophical contributions of women. Volumes also explore ideas drawn from Buddhist, Chinese, Indian, and other philosophical cultures when doing so adds an especially helpful new perspective. By combining scholarly innovation with focused and astute analysis, OPC encourages a deeper understanding of our philosophical past and present.

One of the most innovative features of OPC is its recognition that philosophy bears a rich relation to art, music, literature, religion, science, and other cultural practices. The series speaks to the need for informed interdisciplinary exchanges. Its editors assume that the most difficult and profound philosophical ideas can be made comprehensible to a large audience and that materials not strictly philosophical often bear a significant relevance to philosophy. To this end, each OPC volume includes Reflections. These are short stand-alone essays written by specialists in art, music, literature, theology, science, or cultural studies that reflect on the concept from their own disciplinary perspectives. The goal of these essays is to enliven, enrich, and exemplify the volume's concept and reconsider the boundary between philosophical and extraphilosophical materials. OPC's Reflections display the benefits of using philosophical concepts and distinctions in areas that are not strictly philosophical, and encourage philosophers to move beyond the borders of their discipline as presently conceived.

The volumes of OPC arrive at an auspicious moment. Many philosophers are keen to invigorate the discipline. OPC aims to provoke philosophical imaginations by uncovering the brilliant twists and unforeseen turns of philosophy's past.

Christia Mercer
Gustave M. Berne Professor of Philosophy
Columbia University in the City of New York

Acknowledgments

This book has been a long time in the making—longer than it should have been. Work started in earnest at a conference at Columbia University in March 2014. Because the focus of the book shifted from a greater concentration on seventeenth- and eighteenth-century figures to a longer historical arc, I was unable to include some of the excellent papers that were presented on that occasion. Since the authors, Desmond Hogan, Udo Thiel, and Falk Wunderlich made many helpful contributions to the lively discussions of other papers, their work is also represented in this book, though not in the form I originally intended. Steve Fleming (then at NYU neuroscience, now at UCL) also read a paper on the neuroscience of metacognition at the conference; some of his contributions are part of the basis of Janet Metcalf's *Reflection*. Although a lively participant at the conference, Ed McCann was not able to be a part of this book, and I'm very grateful to Ruth Boeker for stepping in at the last minute with an outstanding essay on Locke. I am also grateful to all the contributors for their patience and, most of all, for their highly original and illuminating essays on our topic. I have learned much from them and I hope that readers will too.

Thanks too to Connie Wang for her careful and efficient editorial work and for preparing the Index.

The series editor, Christia Mercer, has been enormously patient with the fits and starts of this book and has provided a great deal of help as well. Christia had a very good idea for this series and she has worked hard to realize it—especially with this book —and I would like to dedicate it to her in admiration and friendship.

New York City, September 2020

Contributors

CHRISTIAN BARTH is a member of the Philosophy Faculty at the Humboldt University in Berlin. Besides his historical work on, among other topics, "Leibniz and Consciousness," he has contributed to contemporary philosophy of language and mind and is the author of *Objectivity and the Language-Dependence of Thought* (Routledge).

RUTH BOEKER is Assistant Professor in Philosophy at University College Dublin and a member of the UCD Centre for Ethics in Public Life. She is the author of *Locke on Persons and Personal Identity*, 2021.

THERESE SCARPELLI CORY is John and Jean Oesterle Associate Professor of Thomistic Studies, and Director of the History of Philosophy Forum, at the University of Notre Dame. In 2013, she published *Aquinas on Self-Knowledge* (Cambridge University Press).

CHARLY COLEMAN is Associate Professor of History at Columbia University. He is the author of numerous studies of the self, including *The Virtues of Abandon: An Anti-Individualist History of the French Enlightenment*.

DON GARRETT is Silver Professor of Philosophy at New York University. He is the author of *Cognition and Commitment in Hume's Philosophy*, *Hume*, and *Nature and Necessity in Spinoza's Philosophy*.

SACHA GOLOB is Senior Lecturer in Philosophy at King's College, London. His work is on modern French and German philosophy: he is the author of *Heidegger on Concepts, Freedom and Normativity* (Cambridge University Press,

2014), and the editor of the *Cambridge History of Moral Philosophy* (Cambridge University Press, 2017).

ALISON HANSON is a medical resident in the Department of Psychiatry at the Columbia University Irving Medical Center and a member of Rafael Yuste's laboratory.

JARI KAUKUA is a researcher at the University of Jyväskylä in Finland. He has written extensively on Islamic philosophy, including *Self-Awareness in Islamic Philosophy: Avicenna and Beyond* (Cambridge University Press).

PATRICIA KITCHER is Roberta and William Campbell Professor of Humanities and Professor of Philosophy at Columbia University. She has written two books on the relation between Kant's theories of cognition and of the self.

ANTJIE KROG is an Afrikaans poet, writer, and professor in the Faculty of Arts at the University of the Western Cape. Apart from twelve volumes of poetry, she has published *Country of My Skull*, on the South African Truth and Reconciliation Commission; *A Change of Tongue,* about the transformation in South Africa after ten years of democracy, and *Begging to Be Black,* on learning to live within a black majority.

VILI LÄHTEENMÄKI is an Academy of Finland Research Fellow at the University of Helsinki and Senior Lecturer at the University of Oulu. He has written several essays in the history of philosophy of mind, including "Orders of Consciousness and Forms of Reflexivity in Descartes" and "Cudworth on Types of Consciousness."

MARIA H. LOH is Professor of Art History at Hunter College, City University of New York. She has published three books on Titian and on portraiture. Her current research is on representations of the early modern sky.

JANET METCALFE is Professor of Psychology at Columbia University. Her current research is focused on metacognition—how humans know that they know.

GARY OSTERTAG is Affiliated Associate Professor of Philosophy at the City University of New York Graduate Center and Chair of the Philosophy Department at Nassau Community College. He works on philosophy of language and the history of analytic philosophy.

CHRISTOPHER PEACOCKE is Johnsonian Professor of Philosophy at Columbia University and Fellow of the Institute of Philosophy in the School of Advanced Study in the University of London. His most recent book is *The Primacy of Metaphysics.*

PAULIINA REMES is Professor of History of Philosophy at Uppsala University in Sweden. Among other works on ancient philosophy, she has published *Plotinus on Self* (Cambridge University Press).

RAFAEL YUSTE is Professor of Biological Sciences at Columbia. He studies neural circuits in the mammalian cortex and also in cnidarians.

The Self

OXFORD PHILOSOPHICAL CONCEPTS

Introduction

SELF: HISTORY OF A CONCEPT

Patricia Kitcher

1 "I Think Therefore I Am"

No philosophical dictum is better known than Descartes's assertion about the intimate relation between thinking and existing. What remains unknown is how we are to understand the "I" who thinks and exists. This book is about the ways that the concept of an "I" or a "self" has been developed and deployed at different times in the history of western philosophy. It also offers a striking contrast case, the "interconnected" self, who appears in some expressions of African philosophy. Appealing to philosophy to illuminate the concept of a "self" may seem unnecessary. Anyone who can read this book is a self, so why

can we not just tailor a concept to fit what we already know about ourselves? This objection has considerable force and provides a constraint on efforts to fashion a self-concept. Although there is a sense of "self-knowledge" in which it is said to require a lifetime of serious effort to achieve (and which is the topic of another volume in this series), what is at issue here is simply knowing that one is a self.[1] Each of us uses "I" to refer to herself or himself; seemingly, we must have a reasonably good idea of what we are talking about. Further, the most common theme across western philosophical accounts of the self is that selves are self-conscious. As self-conscious selves, readers would have a special vantage point from which they can assess the appropriateness of self-concepts, as opposed to, for example, concepts of "efficient causation" or "space" that are examined in other volumes. Does the proposed self-concept capture what I know to be true of selves by being a self? Although the unique ability of readers to engage with theories of the self constrains philosophers, it also empowers them. Descartes impressed his theory on the minds of his readers so forcefully in part by inviting them to reason along with the meditator, so that when the meditator concludes "*I am, I exist* is necessarily true whenever it is put forward by me or conceived in my mind" (AT VII 24; CSM II 17), so do readers.[2]

2 Who Am I?

Although we can inquire about the reasons for using any concept, the question is pressing for the self-concept. We classify ourselves biologically as "human beings" and morally and politically as "persons"

1 In chapter 1 here, Pauliina Remes discusses texts where self-knowledge as something difficult to achieve has not yet been clearly differentiated from simple self-consciousness of one's own states.

2 References to Descartes's works are to René Descartes, *Œuvres de Descartes*, ed. Charles Adam and Paul Tannery, 12 vols. (Paris: Vrin, 1964–76) (cited as AT). Translations are from *The Philosophical Writings of Descartes*, 2 vols., trans. and ed. John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge: Cambridge University Press, 1984) (cited as CSM).

(bearers of responsibilities and rights), so why do we need still a third concept—"self"—to refer to ourselves? The chapters that follow will present an extensive answer to that question, but the brief answer is evident from Descartes. We think. Tragically, there are cases of human beings who are alive, but who cannot think or who can think no longer. In the latter cases, we could say that the human being had a property, thinking, that she now lacks, but that seems completely inadequate. Relatives and friends are more inclined to feel that, although the bodily shell remains, the individual herself is gone. Often they say that the "person" is gone, so perhaps we could get by just with "human being" and "person." This option is worth exploring, but it has two prima facie weaknesses. Given its "forensic" (Locke) character, the "person" designation seems too broad. Corporations and countries also have legal rights and responsibilities and so are sometimes described as "persons," but they seem to lack the unity of consciousness that has been taken to be the hallmark of human persons in the western tradition. More importantly, characterizing a subject of thought as a "person" blocks fruitful lines of inquiry into the relations between being a thinking self and being a morally and politically responsible agent (see the chapters on Locke and Kant).

This volume presents many reasons for using a self-concept, but it is not exhaustive. Despite their differences, the chapters on western philosophy are largely in agreement on the centrality of unity and of self-consciousness to being a self. A different, but still western, approach to the self-concept might focus on self-control. With that emphasis, a volume might start with Plato's attempt to explain failures of self-control through his tripartite theory of the self, turn to Stoic theories of overcoming desires, and highlight Freud's hypotheses about warring mental systems. Although the capacity to exercise self-control is a major reason for the importance of the self-concept in western philosophy, I did not choose that path, because I think that self-control is a property of being a self rather than what it is to be a self. What a self is—what you most fundamentally are—at least according to

most of the chapters that follow, is a unified self-conscious subject of thoughts.

3 Self-Consciousness, Activity, and Unity

If western philosophy is just a series of footnotes to Plato, then this history of the self-concept seems to start about five hundred years late. As we see in Pauliina Remes's study, although our first author, Plotinus (204–270 CE), begins from several Platonic and Aristotelian assumptions, he raises considerations about the self that were either absent from or peripheral to these important predecessors but which have been dominant themes in later discussions. For example, Aristotle observes that those who perceive that they do,[3] but he does not make the reflexive or self-conscious character of perception central to his account. By contrast, Plotinus tries to analyze the self-reflexive and transparent character of thinking. The Intellect that thinks that it thinks and is aware of its thinking.

Remes notes a second distinctive feature of selves that Plotinus introduces against the background of the ontological realism of Plato and Aristotle. They take the objects of knowledge to precede the knowledge, but do not have an ontological classification of a self. Although Augustine carries this thought much further, Plotinus seems to recognize that the most important aspects of a self are not the sensations that it receives through the body, but its activities. Two of those activities are uncovering the core self and making oneself into something better. For this reason, Remes believes that Plotinus offers a protoconstitutivist account of the self: The metaphor of a statue that must be uncovered suggests a preexisting self, but choosing to emphasize one aspect of oneself rather than another allows selves to constitute themselves, to make themselves what they are.

---

3 Aristotle, *De anima*, trans. J. A. Smith, in *The Basic Works of Aristotle*, ed. Richard McKeon (New York: Random House, 1941), 3.4.425b 12–25.

Augustine (334–430 CE) extends Plotinus's line that it is not what the mind receives through the senses that is crucial to being a mind or self but its activities in relation to them. The mind calls up various materials that are stored in memory, and then acts on them, by connecting them in coherent wholes. We can think of the storehouse of memory as having one unity—some items are stored there whereas others are not—and the connections among those items produced by calling them up and putting them together, cogitating, as a different unity, one produced by thinking. One self will differ from another both in terms of her memories and in terms of what she has done with them. Even in the former case, Augustine encounters himself. Items stored in the vast reaches of his memory are not mere perceptions of things but of things having been experienced by him. Augustine also encounters himself in images of future actions, where he draws on memories and his own creative imagination to plot the way forward. In wandering the spaces of his mind, Augustine connects his past to his future through his cogitation, his current activities of connecting perceptions to each other. For Augustine as for Plotinus, much of the point of exploring the mind's recesses is to make oneself into a better self, but the richness of their discoveries leads many later scholars to focus on what they have found, in particular, to wonder about the nature of this partially self-made self.

4 What Am I?

More than five hundred years after the death of Augustine, the Islamic scholar Ibn Sīnā (Latin: Avicenna, 980–1037) makes a dramatic advance in western philosophical treatments of the self. As with Plotinus and Augustine, Ibn Sīnā's focus is on the self-awareness and activity that is distinctive of thought. Ibn Sīnā takes advantage of his readers being selves and invites them to participate in a *Gedanken* experiment: Imagine that you have just come into existence with no prior bodily sensations or thoughts; just through thinking, you are still

conscious of yourself as something, as existing. Jari Kaukua explains that Ibn Sīnā tries to establish several crucial theses through what was later called the "flying" or "floating" man thought experiment. Self-awareness is both a primitive and a constant feature of thought. The primitive character is brought out through the conditions of the experiment. No prior content, neither sensation nor thought, is needed for self-awareness. Simply by imagining or thinking about yourself as existing, you are conscious of yourself as existing. Although minds can reflect on their thoughts and sensations, selves do not become self-aware through reflection. If a self can affirm itself through a thought or action, then that can only be because the thought or action is (already) the thought or action of that self, prior to reflection. Since this primitive sense of mineness is at least tacitly present in all thought and action, it seems to the subject that she is a constant I across different thoughts and across different activities, for example, thinking and acting. It is the same I that sees an apple, believes it would be good to eat, and reaches for it.

Ibn Sīnā has a further important task for the flying man thought experiment. It is supposed to show that the self is incorporeal: It is a human soul. The reader who performs the thought experiment affirms the existence of herself, without affirming the existence of any body. Kaukua observes that some contemporary readers objected that this inference is too strong. Seemingly, Ibn Sīnā has shown only that a self could exist without a body, not that it does.

Therese Scarpelli Cory explains how Aquinas (1225–1274), despite his lack of a special vocabulary for "selves" or "I"s, synthesizes the insights of Augustine and Ibn Sīnā about self-consciousness and then develops them in ways that run contrary to their original uses. The flying man seemed to have established a self-presence thesis, the thesis that the mind knows itself through its own essence and so may be incorporeal. But Aquinas understands self-consciousness differently. He agrees with Ibn Sīnā that self-consciousness is not a second order or reflective act but analyzes it into two aspects, a disposition to

be self-conscious and particular acts of self-consciousness. As a disposition of the human individual, self-conscious mental activity is not a matter of constituting a new individual (thus pushing back against Augustine); as an act of being conscious of something, it is dependent on that thing (contrary to Ibn Sīnā's claim for the lack of dependence of self-consciousness on anything).

Aquinas also resists Ibn Sīnā's assumption that the self-conscious individual is incorporeal. As Cory notes, his position is not what one might expect from a medieval theologian. The human intellect is an incorporeal power—that is shown by the fact that it is wholly and unerringly conscious of itself in a way that a body could never be—but it is nonetheless a power of an individual human. Aquinas's ontology of a human individual with an incorporeal intellect and a body leads to the odd result that while the self-conscious soul can survive death, the individual person, who is the ensouled human, cannot.

Descartes's dedication of the *Meditations on First Philosophy* presents its twin goals: Prove the existence of God and the immortality of the soul (AT VII 1–2; CSM II 3). Further, as Vili Lähteenmäki observes, Descartes's understanding of what a self is has often been taken to be straightforward. Since thinking establishes the existence of an I, an I must be a thinking thing. From a slightly different angle, Descartes's focus is on certainty in knowledge, and what he can know with absolute certainty is very minimal: I am a thinking thing right now. Putting these together, Descartes's incredibly influential doctrine seems similar to Ibn Sīnā's: A self is, and can be known to be, a thinking, immortal thing, a soul or mind. Lähteenmäki pushes back against these readings by noting that Descartes also identifies the self with the union of the mind and body. Bodily aspects must be included in the self, because most thoughts are about perceptions and feelings and not about mere abstractions. Hence feelings and perceptions are the primary constituents of the human mind; further, they are what inform the individual of what she needs or should avoid. Even if there are excellent reasons to believe that there is an indubitable thinking thing, which would be

a self, and also very strong reasons to identify a self as the union of a mind and a body, it would seem that Descartes cannot have it both ways without falling into vicious ambiguity.

Lähteenmäki argues that, contrary to appearances, Descartes's dual usage is sound, because he is not concerned with the metaphysics of the self—it is the union of a mind and body as opposed to an immaterial mind—but with the question of how subjects use "I" to refer to themselves. They use "I" to refer to a subject of experience, a subject who is aware of her thoughts. Since those thoughts can be about abstractions or about sensations, a self/subject of experience can be either just a mind or the union of a mind and body. Although a self may not understand her metaphysical nature, as a subject of experience, she has what has recently been characterized as a "first-person perspective" on her experience. For example, she knows not only that she is aware of a cat but also how he seems to her, for example, large and ginger-colored.

Lähteenmäki also rejects the traditional reading of Descartes as maintaining that consciousness always involves self-consciousness. Young children might represent a pain but have no representation of the self. Lähteenmäki reads Descartes as offering an explanation for the phenomenology noted by Ibn Sīnā and Aquinas. The vast majority of mental states do include a self-representation, but that is only because they involve attitudes toward contents. Not having been born yesterday, an adult would represent a snarling cat as something to avoid. If Descartes is right, then the self-representation would not be primitive but would be at least partially explicable in terms of attitudes taken toward contents.

5 How Can a Self Represent a Self?

Locke (1632–1704) endorses the cogito argument ("I think, therefore I am") but, noting its limitations, introduces an important new way of understanding selves. Locke agrees with Descartes that in every act of thinking or sensing or doubting an I is conscious of her own existence

with absolute certainty (*Essay* IV.ix.3).[4] As Ruth Boeker explains, Locke sees that this indubitable proof reveals nothing about the nature of the self and that it establishes the existence of an I only in the moment of thought. In a momentary thought, a self is a self to herself—she is conscious of herself as thinking now—but Locke thinks that a self is also conscious of more. The consciousness of the present thought extends back to past thoughts and also forward to possible future pains or pleasures (as Augustine also observed). From the point of view of a subject of experience, "the first personal perspective," selves have duration. Locke's great innovation comes in his controversial account of the enduring self.

Boeker cites Locke's argument that thinking requires some substance in which it inheres: Subjects do not experience thoughts as self-subsistent; in philosophical terminology, they take them to be actions or modes (properties) of some substance to which they are necessarily connected. Since humans have no idea of the nature of that substance, even whether it is corporeal or incorporeal, the sameness of the enduring self cannot be grounded in sameness of substance. Locke invites his readers to consider thought experiments where successive selves might inhabit the same substance or where the same self might change substances. He also considers a hypothetical case where the souls of a prince and a cobbler switch bodies. As Boeker explains, the inability of either selves or philosophical accounts of the self to locate the enduring self in an enduring substance or in an enduring human being raises a major ethical and religious problem for Locke. How can civil or divine rewards and punishments be correctly assigned to the doer of a deed? It is for these reasons, she argues, that Locke takes the concept of a "self" to be that of a "person," which is a "forensic" term indicating moral and legal beings.

4 References to the *Essay* are to John Locke, *An Essay Concerning Human Understanding* (1690), ed. P. H. Nidditch (Oxford: Clarendon Press, 1975); hereafter cited in the text as "*Essay*" followed by book, chapter, and section number.

Boeker shows how well Locke's consciousness-based account of personal identity resolves the legal and religious difficulties stemming from the possible lack of bodily or substantival identity in an enduring self or person. Insofar as the thinker of a present thought or doer of a present deed can extend her consciousness back to earlier acts—can remember doing them—that is how far her personal identity reaches. In that case, no earthly or divine court could punish a person for an action that his own consciousness does not lay at his feet. Although Locke's consciousness-based account of personal identity inspired many subsequent theories,[5] it seems vulnerable to objections. Surely a thief who trips, hits his head on the way out of the bank, and so loses his memory of the robbery is not innocent of the deed!

Boeker explains that Locke's widely discussed *Essay* was immediately subject to penetrating objections and subtle defenses. One of his correspondents, Elizabeth Berkeley Burnet, wondered how the identity of the person could be preserved across gaps in consciousness. If all thinking is conscious, then it must cease in deep sleep. But if Descartes is right that the essence of the mind is to think, then how can a thinking thing cease to think without ceasing to exist? In her *Religious Diary*, Burnet expresses a concern about resurrection. If there is a gap in consciousness, what would be the difference between a person's being resurrected and being recreated? Some of Locke's contemporaries tried to parry the objections. Boeker recounts the careful defense of Catharine Trotter Cockburn against a different objector to gaps in consciousness. Since Locke is right that we do not know the nature of a thing that thinks, how could we know that it could not survive gaps in consciousness? As Boeker notes, Leibniz (1646–1716) tries to resolve the problem in a different way, by denying that all thought is conscious, thus allowing the mind constantly to think or

5 See, for example, Sydney Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca: Cornell University Press, 1963) and Derek Parfit, "Personal Identity," *The Philosophical Review 80* (1971): 3–27.

have perceptions, even if some of the perceptions are so small that they do not reach consciousness.[6]

Christian Barth lays out Leibniz's solution to a different puzzle about the self, a quandary that reverberates through many discussions, including those of Descartes, Locke, and Malebranche. Although Descartes took the idea of the self qua union of the mind and body to be obscure, he thought that humans had a clear and distinct idea of themselves as thinkers just through thinking. (An idea is clear if it can be used to distinguish objects that fit the idea; it is distinct if the different elements in the idea can be captured in a definition.) Locke denies this. He believes that even though selves can know that they think and exist with complete certainty, their only idea of a self is that of an I-know-not-what that thinks. The French philosopher, Nicolas Malebranche, agrees with Locke, explaining that humans have an inner sentiment that informs them that they think, but no clear understanding of what thinking is. As he often tries to do, Leibniz seeks to capture what is right in the opposing positions and thus to produce a more adequate account of what humans are able to understand about themselves as selves.

Barth explains how Leibniz's attempt to mediate between Descartes's view that we have direct knowledge of the self through consciousness and the Lockean and Malebranchean doctrine that we have no knowledge of the nature of the self leads to seeming inconsistencies in his discussions. At points Leibniz claims that we have only a clear, but indistinct idea of the "self," as we have a clear, but indistinct idea of "green." We can instantly recognize ourselves (and green things), but we cannot explain what, exactly, we are singling out. Yet he also claims that it is by reflection on the nature of the mind or soul that humans come to have metaphysical notions such as those of "being," "substance," "unity," and "action." How can both of these claims be correct?

---

6 Gottfried Wilhelm Leibniz, *New Essays on Human Understanding*, trans. Peter Remnant and Jonathan Bennett, abridged ed. (Cambridge: Cambridge University Press, 1982), e.g., 54.

Barth argues that we can understand Leibniz's view only by following his distinction between an "idea" and a "concept" and his distinction between two senses of "reflection." A "concept" is a mental state that occurs at a moment, for example, a perception of a cat. By contrast, an "idea" is the disposition to have concepts. Lacking that disposition, one could not have a cat-perception however many cats walk by. So Leibniz maintains that humans are able to be conscious of themselves in thought, because they have a disposition to form the representation of a "self" while thinking. As with the concept "green," this concept allows them to distinguish a self from other things. Still, Leibniz agrees with Locke and Malebranche, that the perception of a self is obscure. We do not know what properties it has and so do not know the ideas that are involved in the self-concept.

How humans move from the obscure perception of the self to a rich metaphysical understanding of the nature of selves is by two types of reflection. First, because humans have inner sentiments of themselves, they can reflect on what they are perceiving. Through reflection on her thoughts, a self might perceive that she is, for example, thinking about an evil demon deceiving her. Further reflection might disclose that she also thinks about basic arithmetic truths and thus that it is the same self that has both thoughts. With these materials from the first sort of reflection—being aware of what you are thinking—the self can move on with the help of innate metaphysical ideas to try to determine her nature. For example, since she has two different thoughts or properties, then she must be a continuing substance who is the bearer of different properties. Since she thinks both these thoughts, she is not a passive substance, but an active one. Through this second form of disciplined reflection, guided by innate ideas of metaphysics, a self can use what she reflects on at the first level to form an ever more distinct idea of her nature as a self. Hence, Locke can be right that a self's original perception (or "idea" in his sense) of a self is obscure, and Descartes can also be right that through thinking (combined with metaphysical reflection,

which is what the meditator seems to engage in), a subject of thought can come to have a clear and distinct idea of his nature as a thinking thing. For all its elegance, Leibniz's resolution of the paradox of only being conscious of the self in thought as an "I-know-not-what" and philosophers' attempts to give a rich metaphysical account of the self depends on many of his assumptions about innate ideas, and those assumptions are highly controversial.

Hume (1711–1776) rejects any doctrine of innate ideas in the human mind and, in particular, claims about an innate idea of the "self." As Don Garrett explains, Hume's powerful arguments against his predecessors' views about the nature and origin of the idea of the "self" led his contemporaries and later readers to attribute an extreme position to him: There is no self or at least no idea of a "self." Although Garrett argues against this traditional understanding of Hume's views, he concedes that Hume did not help his readers, because he did not develop his positive theory as much as was needed.

Garrett offers a number of considerations to show that Hume did not so much reject a self-concept as develop a novel account of it. Since he denied innate ideas, the only idea of a "self" that humans could have is one produced through sensory impressions. For Hume, ideas are copies of impressions, that is, fainter versions of sensory impressions. So he considers his impressions and tacitly invites readers to do the same. This exercise reveals that there is no constant impression of a "self," but only a sequence of perceptions (impressions and also ideas copied from them). Those philosophers who claim that in consciousness humans are directly and indubitably aware of a self that endures through time and is the bearer of mental states (a substance) are simply mistaken (THN 1.4.6.2, 1.4.5.3–5).[7]

7 References to David Hume, *Treatise of Human Nature*, ed. P. H. Nidditch (Oxford: Oxford University Press, 1978), are indicated in the text by THN and the book, part, section, and paragraph.

Garrett lays out Hume's positive view: If ideas come from impressions and what humans are conscious of is a rapidly changing sequence of perceptions, then the idea of a "self" would be that of a system of perceptions linked by relations of cause and effect (THN 1.4.6.19). By its nature, the complex idea of an individual "self," what Garrett calls a "basic individual idea,"[8] cannot be fixed or determinate. New perceptions are added, and not all perceptions are always recalled. As with other individual ideas, the idea of a "self" can be made general, because it can refer indifferently to any bundle of perceptions that is interrelated as the individual person's perceptions are interrelated. In Garrett's view, Hume does not deny that humans have a general idea of a "person" or "self," but only that they have the idea that his predecessors attributed to them, namely, that of a simple enduring substance that is the bearer of perceptions. Their idea is actually that of a complex sequence of perceptions that stand in various relations which lead us to associate them together as perceptions of one "self."

Garrett's dislodging of the "no self" or "no idea of a self" interpretations of Hume allows him to answer questions that have plagued those readings. I mention just three. Many commentators claim to find a contradiction between book 1 of the *Treatise*, where Hume is said to deny any idea of a "self," and book 2, which offers lengthy accounts of pride and shame, passions that have the "self" as their object. Similarly, Hume's account of the concern that humans have for themselves and their futures would conflict with the denial of any idea of the "self." Such a denial would also be inconsistent with Hume's appeal to each person's idea of her "self" to explain the mechanism of sympathy, the emotion that he takes to be an original source of moral judgment. If Garrett is right that Hume has a positive idea of a "self," then these alleged and embarrassing inconsistencies disappear.

8 Don Garrett, chapter 7 here, p. 185.

6 Must Humans Have a Unified Self?

Hume described the *Treatise of Human Nature* as providing a "science of man" that would be built on the same foundation of experience and observation that supports the science of nonhuman nature.[9] In his ethical writings and discussion of natural religion, he also recoiled from what he took to be the excesses of religious mysticism.[10] In "Diderot, the Self, and the Science of Dreaming," Charly Coleman examines some of the countercurrents that challenged the enlightenment's commitment to the idea of an active, self-conscious, unified "self." As he notes, the questioning came from two sides, from a scientific materialistic account of human nature in terms of neurons and causal forces that had no place for a unified or uniting self and from a heretical religious mysticism that focused on aberrant experiences, such as dreaming and distraction, that seemed not to be possessed by a self-conscious self.

Coleman's anti-enlightenment texts are principally entries from the greatest collective expression of the enlightenment, the French Encyclopedia (1751–1772)! Diderot (1713–1784) rejects many of the claims for "altered states" of consciousness, but dreams and daydreams are familiar phenomena. Descartes worried about dreams as well as demons and, as we have seen, dreamless sleep posed a widely recognized threat to the continued existence and identity of a self. Diderot's encyclopedia articles on dreaming raise a different issue. The self-conscious self seems to go missing in dreaming and in some states of distraction. There is no sense of ownership of the experiences and no possibility of applying critical reflection to them. Since these phenomena are undeniable, they suggest that a (constantly) "self-conscious self" may be a philosophical ideal rather than an existing thing. Coleman takes Diderot's entries on eclecticism and delightful repose (*délicieux*),

9 Introduction to THN.

10 See, e.g., David Hume, *An Inquiry Concerning the Principles of Morals*, ed. J. B. Schneewind (Indianapolis: Hackett, 1983), sec. 9, and his *Dialogues Concerning Natural Religion*, ed. Richard H. Popkin (Indianapolis: Hackett, 1982), pt. 10.

as well as those on dreaming, to be evidence that he recognized that a human being in possession of a self could become dispossessed or alienated from that self.

Kant (1724–1804) considered the question of the unity of the self in terms of his distinctive philosophical approach. His focus was not on whether the unity of the self and self-consciousness could be disrupted or broken, but on why it was necessary to have a concept of a "self" which represents a self as unified and self-conscious. His answer is that any creature capable of cognition must have such a concept, because the use of that concept is a necessary condition for the possibility of cognition. Unless cognizers use the concept of a "self-conscious unified self"—understand themselves as such selves—they would be incapable of the simplest cognition, that is, of judging that something is some way.

Kant maintains that cognition through the senses requires two types of mental faculties and two types of mental states or representations: a passive faculty of sensibility that receives representations (which he calls "intuitions") and an active faculty of understanding that combines intuitions in conceptual representations or "concepts." Kant explains that concepts rest on "functions" which organize different intuitive elements in a single representation, for example, the tail, ears, muzzle, and so on as parts of one whole dogish representation, or which collect different whole (dogish) representations under a single conceptual representation, "dog" (A68/B93, B130).[11] Even to judge that a dog is brown, a cognizer needs to receive information through her senses and then order that perceptual information under general representations or concepts. A central goal of the *Critique of Pure Reason* is to demonstrate that, although the concepts through which sensory cognition is possible are acquired through sensory experience, they are built on templates supplied by the "categories," i.e., a priori concepts that are

11 References to Immanuel Kant, *Critique of Pure Reason: Unified Edition*, trans. Werner Pluhar (Indianapolis: Hackett, 1996), are given in the text with the usual A/B pagination.

not acquired through sensory experience. For example, one category is "substance," a thing with properties, and the concept "dog" is the concept of a substance. If cognition is not just a matter of receiving sensory representations but also of combining them in accord with a priori concepts supplied by the subject herself, then cognition is a reflection of active powers of the mind and not just of the sensory materials that it receives.

Kant's doctrine of the unity of the self is central to his account of cognition, because that unity is the fulcrum of his argument that cognition requires combination in accord with a priori concepts. He argues first that cognition requires that a subject compare and connect representations (A97) that she understands as belonging to a single self (A116, A117a) or, as he puts it, that belong to the same "I think" (B131–32). He then offers extensive (and controversial) arguments intended to show that a subject can understand different representations as belonging to the same self if and only if she combines representations in accordance with the categorial concepts (A108, A111–12, B134, B143). A surprising result follows: Although cognizers must represent themselves as unified selves for cognition to be possible, they can do so only by combining representations in order to cognize objects. For that reason, cognizers' consciousness of their own unity and their cognition of objects must always occur together. The representation of a unified subject—an "I think"—is similar to the categories in being a priori. It cannot be received through any sense, but arises through the activity of comparing and connecting representations when cognizing objects (B1–2). If Kant is right, then it follows that Leibniz's adherents were mistaken in maintaining that the representation of a "simple, single, unified self" is the result of applying the principles of metaphysics to an I they are conscious of in thinking: Is a thinking I a substance or an accident, simple or complex, momentary or continuing (e.g., A404)? It would also follow that Hume erred in looking for a self-representation in sensory impressions rather than in the activity of thinking.

When Kant turned to ethics, he used the same distinctive approach that he applied to epistemology. Instead of asking how cognition is possible, he asks, "how is morality possible?" His well-known theory is that morality is possible only if agents can act from the categorical imperative: "Act only according to that maxim that you can at the same time will that it become a universal law" (4.421).[12] Kant believed until late in life that the crucial feature of selves as opposed to mere animals—their freedom to think and to act in ways that were not completely guided by what they received from the senses—was the same in the cases of cognition and morality. As he came to see the implications of his moral theory more fully, however, he realized that these freedoms come apart: The freedom required for morality is not just liberation from domination by the senses, but a positive capacity to act on the basis of moral considerations.

As I argue in chapter 9, Kant's investigations into the necessary conditions for cognition and morality lead to a strange situation: The thinking I and the acting I can come apart. It is necessary for the possibility of morality that, however badly someone has behaved, she can do the right thing in the very next action (4.454, 5.79, 6.36, 439). Yet Kant also thinks that if a moral subject, a person, changes dramatically from, say, always acting selfishly to following the principles of morality, then she has become a new person and so is no longer responsible for the deeds of the "old" person. As he realizes, the new person would, however, be the same thinking self as the old person, because she would understand her current judgments as belonging to the same thinking self as earlier judgments. In chapter 12, Antjie Krog recounts a case where this possibility may have been realized.

---

12 References to Kant's works other than the *Critique of Pure Reason* are given in the text and are to the volume and page of Immanuel Kant, *Kants gesammelte Schriften*, Akademie Ausgabe, ed. Koeniglichen Preussischen Akademie der Wissenschaften, 29 vols. (Berlin: Walter de Gruyter and predecessors, 1900–). The translation is from Immanuel Kant, *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor and Jens Timmermann (Cambridge: Cambridge University Press, 2011).

7 Recent Approaches to the Self and Self-Consciousness

Kant's German idealist successors greatly elaborated the doctrines that the self is active in producing cognition and free from the limits of sensory experience in deciding how to act. As Sacha Golob argues, the Phenomenological tradition made an important break from those projects by introducing a new method for studying the self and self-consciousness. The historical movement that has included Heidegger (1889–1976), Sartre (1905–1980), and Merleau-Ponty (1908–1961) has argued that Fichte (1762–1814) and Hegel (1770–1832), like Kant, took an overly intellectualist approach to the self and, especially, to self-awareness. The "Phenomenologists" urged that rather than consider self-consciousness in relation to the use of concepts and ideas, philosophers should study what is available to the self in self-awareness. Golob notes that although different members of this tradition had different foci, they were agreed that their predecessors erred in not trying to uncover primitive forms of self-awareness, the forms that make the sophisticated forms they did consider possible.

As we have seen, Descartes thought that what the "I" most fundamentally refers to is a subject of experience, a subject with a first-person perspective on what her experience is like. The Phenomenologists' critique is that Descartes did not explore this perspective sufficiently, but moved immediately to thought, to *cogito*. Had he stayed within the first-person perspective, he would have discovered a number of features that cannot be captured in concepts and so cannot be captured by science. Golob catalogs some of the features that Phenomenologists take themselves to have found.

Sartre introduces the notion of "nonpositional" consciousness to oppose the idea that self-consciousness must involve consciousness of a self as something separate from conscious awareness. His example is that of counting the cigarettes left in a pack. Although the counter could say what she was doing if asked, that does not imply that she

was (previously) conscious of herself as opposed to simply being conscious in counting. In Sartre's view, the very fact that the counting is conscious implies that it includes self-awareness, because the smoker is aware of how things are with her, counting the first cigarette and so forth. Where Descartes and Kant err, Sartre believes, is in thinking that this primitive type of self-awareness requires either concepts or reflection on what one is doing.

Heidegger stresses that the subject's awareness of her temporal and spatial experience does not track scientists' representation of these phenomena: From the point of view of the subject, the friend across the street is closer than the glasses on your nose. Beyond the distinctive subjective awareness of space and time, Heidegger highlights the primitive type of self-awareness that is evident in everyday life. A professor entering a lecture hall is aware of the lectern not in the sense that she makes a judgment about it—"lo, a lectern"—but that she "knows how" to behave in relation to it. She might place her notes on it, stand behind it, and raise the microphone to her height. Although she has the concepts "lectern," and so on, there are a myriad of appropriate movements that she makes without having any concepts to characterize them. Yet, she not only makes the appropriate movements and gestures, but is also aware of herself as doing so. So she is self-conscious, but not in the sense of reflecting on what she is doing (as Heidegger also thinks that Descartes and Kant maintain), but simply as acting in particular ways.

With Merleau-Ponty, we return to an issue that arose with Ibn Sīnā and Descartes and that has been widely discussed in accounts of the self, the relation of the body to the self and its representation. Merleau-Ponty argues that a primitive form of self-awareness involves a representation of the "body." Building on Heidegger's notion of "knowing how," he notes that action requires some sort of representation of where the agent's body is. A tired carpenter might hit her thumb instead of the nail, but carpenters mostly avoid this sort of mistake. To do so, they need to have not just an accurate representation of where the nail is in relation to them (an egocentric map) but also a representation

of how they are moving in relation to the nail and the hammer. Again, Merleau-Ponty insists that the requisite self-awareness is independent of concepts and judgments. It is, rather, a prerequisite for action.

As Christopher Peacocke presents progress on the self-concept in the twentieth and twenty-first centuries, he acknowledges the debt of nonphenomenologists to that tradition. In particular, significant post–1945 work on the self-concept has stressed the phenomenal character of mental states such as color perception and emotions, characters that are available only from the first-person perspective. Peacocke notes that this work on the nature of first-person experience has been carried forward in the new (in the twentieth century) field of philosophy of mind. Peacocke credits another recent field—philosophy of language—with answering a basic question raised right at the beginning: How do subjects use "I" to refer to themselves (as selves)? The "fundamental reference rule" for "I" is that in any mental state or linguistic utterance that includes the content "I," "I" refers to the owner of the state or the maker of the utterance.[13] As Peacocke explains, although the fundamental reference rule does not answer questions about the nature of selves and so give a full account of the "sense" of "I," it does bring out unusual features of the I-concept. "I" is indexical, indexed to context: It changes its reference on different occasions of thought or utterance. If Peacocke thinks, "I am trying to understand the nature of selves" and if you have that same thought, "I" has a different referent in those two thoughts.

Recent work on the self-concept has drawn attention to another unusual feature of "I" thoughts and utterances. Wittgenstein (1889–1951) contrasted two uses of "I," one as object and one as subject. In the first sort of use, someone could assert, "I am getting a sunburn," by seeing a red shoulder in the mirror. If the shoulder is not hers but a passerby's, she is mistaken. By contrast, if someone thinks "I am

13 Christopher Peacocke, chapter 11 here, p. 288.

trying to understand the properties of an isosceles triangle," then she might be wrong, because what she has in mind is actually a scalene triangle. What she could not be mistaken about is that she is the subject of the thinking. As Peacocke notes, Sydney Shoemaker offered a precise account of Wittgenstein's distinction. When "I" is used to designate an object, there can be an error due to misidentification of that object (the wrong person's shoulder). In the second case, however, when "I" is used as "subject," it is "immune to error by misidentification" (IEM). That does not amount to Cartesian certainty about the thought, since the subject could be confused about content, but the IEM phenomenon may partly explain why earlier philosophers believed that thinking was certain in a way that perceiving "external objects" could never be. One large question about this phenomenon is how far it extends. Seemingly "I-thoughts" about action, such as G. E. M. Anscombe's classic example "I am pumping water," are also IEM;[14] as Peacocke notes, Gareth Evans argued that claims about the disposition of one's body are also IEM.

Peacocke argues that even when the epistemologically peculiar feature of I-thoughts that they are IEM is added to the fundamental reference rule for "I," we still do not have the full sense of "I." That requires a metaphysical account of the nature of the self. As he documents, the long twentieth century returned to some of puzzles and options offered by earlier theorists, but also added scientifically informed new possibilities. Is there no self (as many took Hume to maintain)? Must a self be embodied (as Merleau-Ponty seems to suggest) or does an updated version of Locke's thought experiment—switching brains instead of minds—show that having the same body is not essential to being the same self? Does an updated version of Ibn Sīnā's flying man thought experiment, where someone is placed in a sensory deprivation tank and then deprived of memories, yet

14 G. E. M. Anscombe, *Intention* (Ithaca, NY: Cornell University Press, 1969), secs. 8 and 23.

still seems to retain her ability to self-ascribe thoughts, imply that interaction with objects and memory are also inessential to being a self? If a self must understand itself in relation to others, is it part of the concept of a "self" that a self belongs to a social world (as Sartre and Hegel before him seem to think)? A new option that Peacocke commends is understanding a self in terms of the apparatus that integrates perceptual materials in such a way as to produce egocentric perception even if the self is not part of that representation. By starting with this primitive sort of awareness, it might be possible to analyze the different layers of capacities that are part of the sense of a "self." Peacocke's strategy of starting with the most primitive types of representation associated with a self (subjects that may exist at "Degree 0") parallels the concern Phenomenologists have about beginning with primitive forms of self-awareness. Where his approach and that of the figures he presents often differs from that of Phenomenologists is in its willingness to draw on the sciences of psychology and neuroscience in understanding the nature of selves; as Golob explains, Phenomenology takes its subject to be utterly divorced from the methods and results of natural science.

8 Interconnected Selves in the African Tradition

As Peacocke explains, although many branches of philosophy can illuminate how the self-concept works, the account will be incomplete until the metaphysics of the self is filled in. Despite its emphasis on unity and self-consciousness, the western tradition has also noted that humans represent themselves as one self among many. It does not, however, go as far as the African tradition that Antjie Krog presents in understanding selves and living beings more generally as parts of an interconnected whole. Putting the point in Augustinean terms, in this African tradition, the unformed self constitutes itself through realizing its interconnectedness with others. Krog explains how the metaphysics of interconnected selves undergirds very different conceptions

of crucial ethical notions, such as recognizing the humanity of others and forgiving them for misdeeds.

Krog brings out the distinctive features of a traditional African self-concept by considering a heinous and controversial case from the work of the Truth and Reconciliation Commission. Eugène de Kock had been in charge of a unit of the (apartheid) South African government whose remit was to kidnap, kill, and torture antigovernment activists. He confessed and asked for amnesty. He was sentenced to twenty years in jail. Krog's account begins as he is about to be released. During his time in prison he had spoken to relatives of the many people whom his unit had killed or brutalized. Although he could receive no reduction in his sentence, telling his victim's relatives about what had happened to their loved ones brought some closure to their suffering. Was Eugène de Kock then the new man of the Gospels (and Kant), a man who had changed so much that his past crimes were somehow no longer his? Krog explains that this is not how some traditions within African philosophy would characterize the situation. Understanding one's self as part of an interconnected whole implies that bad behavior detracts from the humanity of all. Taking restorative measures does not lead merely to some people forgiving others, but to repairing the humanity in the victims as well as that in the perpetrators. As Krog argues, a western approach to the self makes forgiveness seem superhuman in this sort of case. With a different understanding of the "self," however, releasing Kock after he had served his sentence was not a matter of forgiving the unforgivable but of rehumanizing someone who had lost his humanity, and thereby detracted from the humanity of others. Not everyone in South Africa agreed with this level of reconciliation, but Krog thinks that we should understand those who did not by lionizing their individual virtue but by recognizing their distinctive metaphysics of the self—a metaphysics expressed not just by African philosophers but also by some of the African songs and poetry that she presents.

9 Reflections on the Self in Painting and Poetry

Self-portraiture is a familiar genre. Yet, as Maria Loh explains in a reflection on Caravaggio, it is full of paradoxes. A self-portrait is recognized as such by depicting the artist's body—usually the face—not the self. To do that, the artist must see himself not as he normally sees himself but as others see him, in a frozen moment, not even as a moving body but as a still life. In the case study Loh presents, Caravaggio seeks to make his first-person perspective as a self visible to others by presenting a likeness of his face as belonging to the severed head of a historical (or mythical) individual (1610). The well-known and grotesque facts of the story of David and Goliath enable Caravaggio to present his face, and so his self, in a cri de coeur of utter loss. Loh notes that an earlier self-portrait, with Caravaggio as Narcissus (1599), seems to be a study of the paradoxes of the self-portrait. Caravaggio the artist paints, and so looks at, as an object, Caravaggio, who, as a subject, is gazing at the watery reflection of himself as an object.

Where Loh uses self-portraiture to illuminate tensions between the first-personal and objective points of view, Gary Ostertag reflects on a poem that helps to bring out what a strange sort of being a self is. Elizabeth Bishop's "In the Waiting Room" (2011) describes the young Elizabeth's reaction to recognizing that she is a "self," an "I." The recognition comes through a momentary apparent merging of her identity (while reading a magazine article about cannibals that frightens her) with that of her aunt, who is undergoing a dental procedure. There is a cry of pain, "oh!" that she thinks comes from her aunt in the dentist's chair but is made with her own voice. To move beyond the "we," the sense that she and her aunt are somehow falling (perhaps together), she focuses on a unique fact about herself. It is soon to be her seventh birthday. But this just leads to more bewilderment. The recognition that she is an "I," a "self," leads her ask why she should be a self, one (particular) self among other selves, and to wonder, as Descartes did, just what a self is.

10 Reflections on Sciences of the Self

The two fairly constant themes of self-conscious selves and the unity of the self in western philosophy have been picked up by contemporary researchers who aim to bring the resources of science to bear on thinking about selves. In cognitive psychology, the difficult notion of self-consciousness has been approached through "metacognition," that is, cognition of cognition. Psychologists have investigated the frequent philosophical assumption that self-consciousness is restricted to humans in cleverly designed experiments on metacognition in monkeys and other animals. Since animals and preverbal children cannot report on their thoughts, researchers have had to find indirect ways of determining if they know what they know. Janet Metcalf presents two (similar) methods that have been used to reveal monkeys expressing their confidence in what they know through whether they accept a passable reward or risk losing that reward for a more desirable one. Besides behavioral evidence about the range of self-consciousness, Metcalf also presents work in brain imagery that aims to find the location of capacities such as metacognition that are necessary for being a self. Brodmann's Area 10 of the prefrontal cortex seems to be activated in many tasks involving metacognition and is progressively more developed in animals capable of metacognition. As Metcalf notes, however, it is too early to tell whether that area should be considered the seat of the "self" in the brain.

Neuroscientists Alison Hanson and Raphael Yuste present recent work that focuses on the unity of a self, in particular on living systems that are able to draw a boundary between themselves and what is not themselves. Hanson and Yuste explicitly reject the thesis that a self must be self-conscious and propose instead to examine how even organisms lacking a nervous system can distinguish what is part of themselves as opposed to what belongs to the "external world." They suggest that, for example, a bacterium models itself and the environment on which it must act to survive. As they observe, the presence of a

primitive self in the absence of neural tissue raises the question of what the neural system—assumed to be the seat of self-consciousness and so the self—contributes to being a self. One possibility is that neurons are not the self, but rather make it possible for different parts of a self to communicate quickly with each other. Perhaps, though, matters are different in larger, more complex organisms. For humans and others in this category, Hanson and Yuste consider the possibility that the spontaneous (i.e., not triggered by external stimuli) synchronous firing of neurons that is common to resting states and to self-related mental activity such as memory and spontaneous thought, what they call the "default mode network" (DMN), may be key to being a unified self;[15] its location seems to make it apt for integrating information from different brain locations. Like Metcalf, they are skeptical of identifying the DMN or any one structure as the seat of the "self." For these scientists, as for some contemporary philosophers, the goal is to illuminate different aspects of the self by considering primitive forms and then trying to determine how these are built upon in the evolutionary history of complex organisms.

15 Alison Hanson and Rafael Yuste, Reflection here, pp. 309–10.